Analysis of Health Utility Data

Eleanor Pullenayegum
Asst Professor
Clin Epi & Biostats
McMaster University

A Story

- Analysis of EQ-5D data
- Estimate decrease in health utility upon occurrence of diabetes-related adverse events
- Results will inform the Ontario Diabetes Economic Model
- Plan was to use the same methods as were used in a similar UK study

Objectives

- To describe the features of health utility data
- To outline how it is typically analysed
- To outline how it should be analysed

- Main point: communication breakdowns have led to some strange analyses in this field!

What is a Health Utility, and what is it used for?

Health utility data – what is it used for?

Economic analysis

- Calculating QALYs
- Calculate difference in mean QALYs between two treatment options

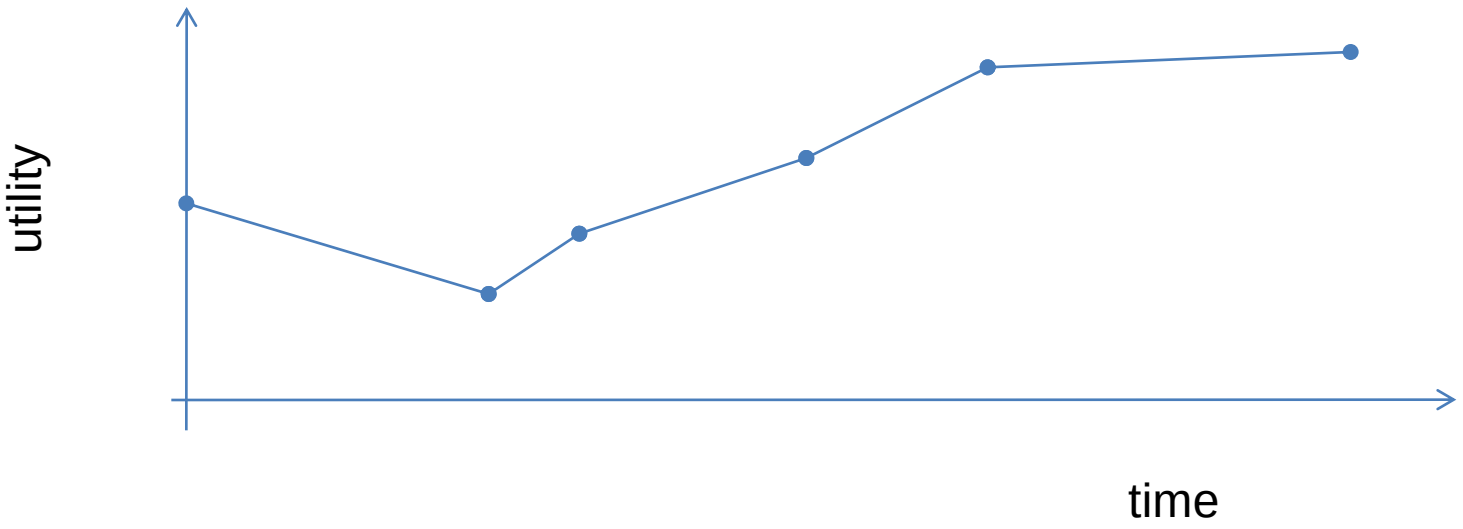
Measuring quality of life

- Sometimes, health utilities are used simply as a measure of quality of life

Utilities & QALYs

- QALYs accrued over time 0 to t given by

$$\int_0^t \text{utility}(s)ds$$



Utilities

- A utility is a quality weight used to calculate QALYs
- A utility of 1 represents full health
- A utility of 0 represents a state equivalent to death
- Utilities can be negative

Health-related quality of life

HRQoL

- Measures quality of life
- Abstract construct (like IQ)
- Unbounded
- Need not have interval or ratio properties

In what contexts do we collect utilities?

- RCTs
  - Secondary outcome
  - May be used to inform a cost-effectiveness analysis
- Observational studies
  - Can be cross-sectional, longitudinal
  - Need to adjust for confounders
  - If used for economic analysis, often in the context of a complex economic model

How are health utilities measured?

- Usually *indirectly*
- Generic instruments
  - EQ-5D
  - HUI
  - SF6D
- Disease-specific instruments
- Based on the response to the questionnaire, there is a scoring algorithm to get a utility

Your own health state today

By placing a tick in one box in each group below, please indicate which statement best describes your own health state today.

Do not tick more than one box in each group.

Mobility
I have no problems in walking about
I have some problems in walking about
I am confined to bed

Self-care
I have no problems with self-care
I have some problems washing and dressing myself
I am unable to wash and dress myself

Usual activities (eg. work, study, housework, family or leisure activities)
I have no problems with performing my usual activities
I have some problems with performing my usual activities
I am unable to perform my usual activities

Pain/discomfort
I have no pain or discomfort
I have moderate pain or discomfort
I have extreme pain or discomfort

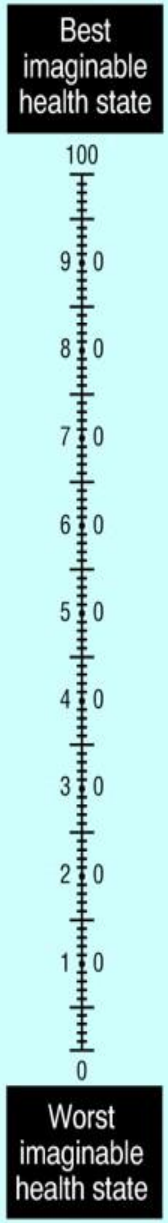
Anxiety/depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed

Your own health state today

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is.

Your own health state today



EQ-5D UK scoring algorithm

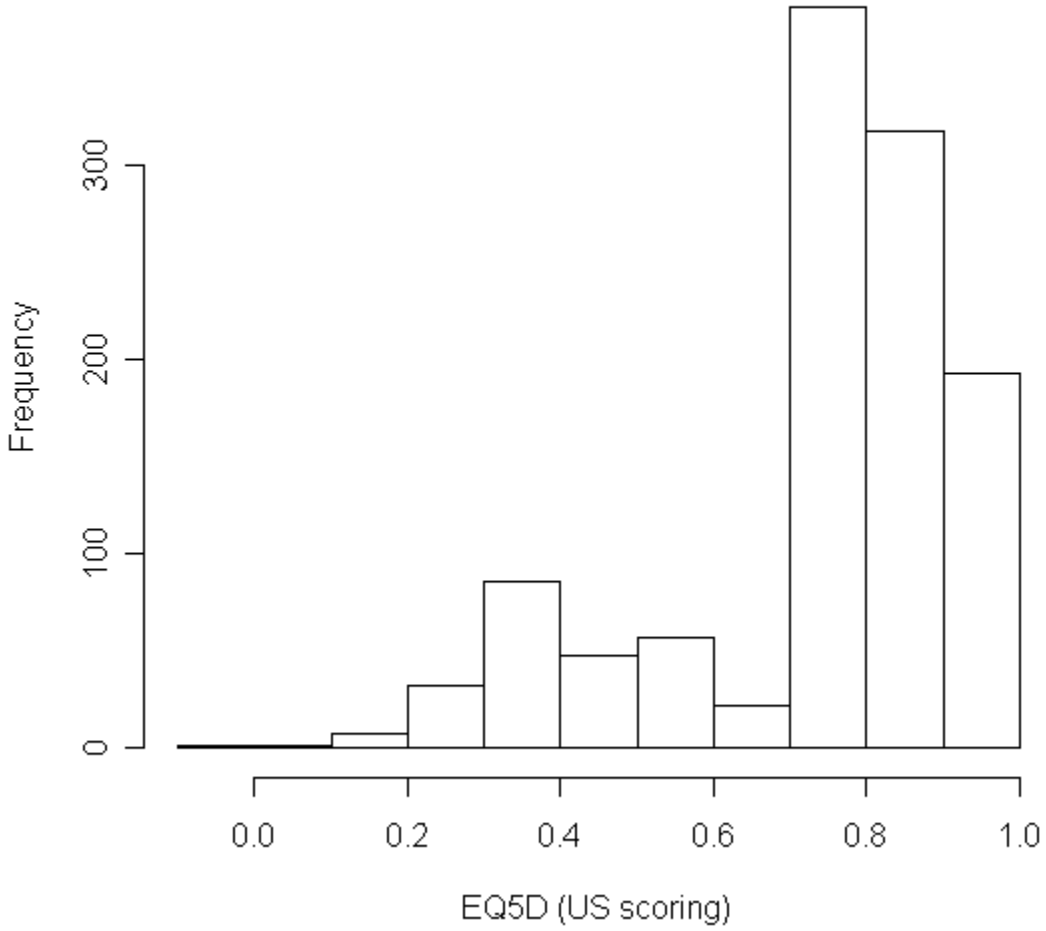
Full health	**1.000**
Constant term (for any dysfunction state)	-0.081
Mobility level 2	-0.069
Mobility level 3	-0.314
Self-care level 2	-0.104
Self-care level 3	-0.214
Usual activities level 2	-0.036
Usual activities level 3	-0.094
Pain/discomfort level 2	-0.123
Pain/discomfort level 3	-0.386
Anxiety/depression level 2	-0.071
Anxiety/depression level 3	-0.236
N3 (level 3 occurs for at least one dimension)	-0.269

How are Utilities Analysed?

Example: The Ontario Diabetes Economic Model

- Current estimates in the model are from the UK
- Wanted to update the model with Canadian data
- Cross-sectional data from 1141 Canadians
  - Health utilities (captured through the EQ5D)
  - Adverse events (Stroke/ MI/ Kidney Failure/ Amputation)
  - Confounding variables (Age, gender etc.)
- Estimate adjusted mean difference in utility amongst those with and without each adverse event.

Health utility data



Distribution is

- Non-Normal
- Often bi-modal
- Bounded below
- Bounded above at 1
- A number of patients achieve the upper bound of 1

Point #1:
Utility Data Has a Strange Distribution

Analysis

- Two-part models
- Latent class models
- Beta models
- Tobit models
- CLAD models
- Linear regression

Tobit models

- Assume observed utility has been censored at 1
- True utility follows a Normal distribution

$$y^* \sim N(X\beta^*, \sigma^2)$$

$$y = 1 \text{ if } y^* > 1$$

$$y = y^* \text{ otherwise}$$

y* =True utility

y = Observed utility

Tobit Model

CLAD models

- Censored Least Absolute Deviations
- CLAD models minimize

$$\sum_{i} |Y_i - \min(X_i\beta, 1)|$$

- Model the median, not the mean

Censoring?

- Is utility data censored at 1?
- Could it be possible to accrue more than one QALY in a year?
- E.g. EQ5D and HUI scoring algorithms assume that 1 represents “full health”.

- Counter-argument: there exist supranormal health states that should have a utility larger than 1

Coarse measurement

Mobility

I have no problems in walking about ❑

I have some problems in walking about ❑

I am confined to bed ❑

Self-Care

I have no problems with self-care ❑

I have some problems washing or dressing myself ❑

I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑

I have some problems with performing my usual activities ❑

I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑

I have moderate pain or discomfort ❑

I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑

I am moderately anxious or depressed ❑

I am extremely anxious or depressed ❑

Fewer people would have scores of 1 if there were more response levels

Does it matter?

- Does assuming that utilities are censored above at 1 lead to bias?
- Simulation study:
  - Took samples of 50, 100, 200, 500 and 1141 individuals from our study
  - For each SS, sampled 1000 datasets from the empirical distribution
  - Used Tobit & CLAD to get estimates of the difference in utility between those using and not using insulin

Simulation Results

Sample Size	True Effect	Bias (bias/se)	
		Tobit	CLAD
50	-0.066	-0.00252 (-1.17)	-0.00108 (-0.72)
100	-0.100	-0.00327 (-2.95)	0.05721 (64.10)
200	-0.055	-0.01034 (-10.23)	0.03497 (34.14)
500	-0.049	-0.00482 (-7.09)	0.03433 (54.93)
1141	-0.073	-0.00587 (-12.76)	0.04225 (85.67)

Simulation Results: Empirical Coverage Probabilities

Sample Size	Tobit	CLAD
50	0.939	0.660
100	0.924	0.217
200	0.942	0.725
500	0.943	0.272
1141	0.921	0.029

Point #2:

Tobit & CLAD models are not appropriate for utility data when estimates will inform an economic model

Why did people start using Tobit/CLAD models for utilities?

- Often cite Austin et al
  - Austin PC. A comparison of methods for analyzing health-related quality-of-life measures. Value in Health 2002;5(4):329-337.
  - Austin PC, Escobar M, Kopec JA. The use of the Tobit model for analyzing measures of health status. Quality of Life Research 2000;9:901-910
- These studies were looking at predictors of quality-of-life when captured using scales (e.g. HUI)
- Not looking at utilities

Warning #1

- What **we** write in **other peoples'** literature sometimes gets mis-interpreted

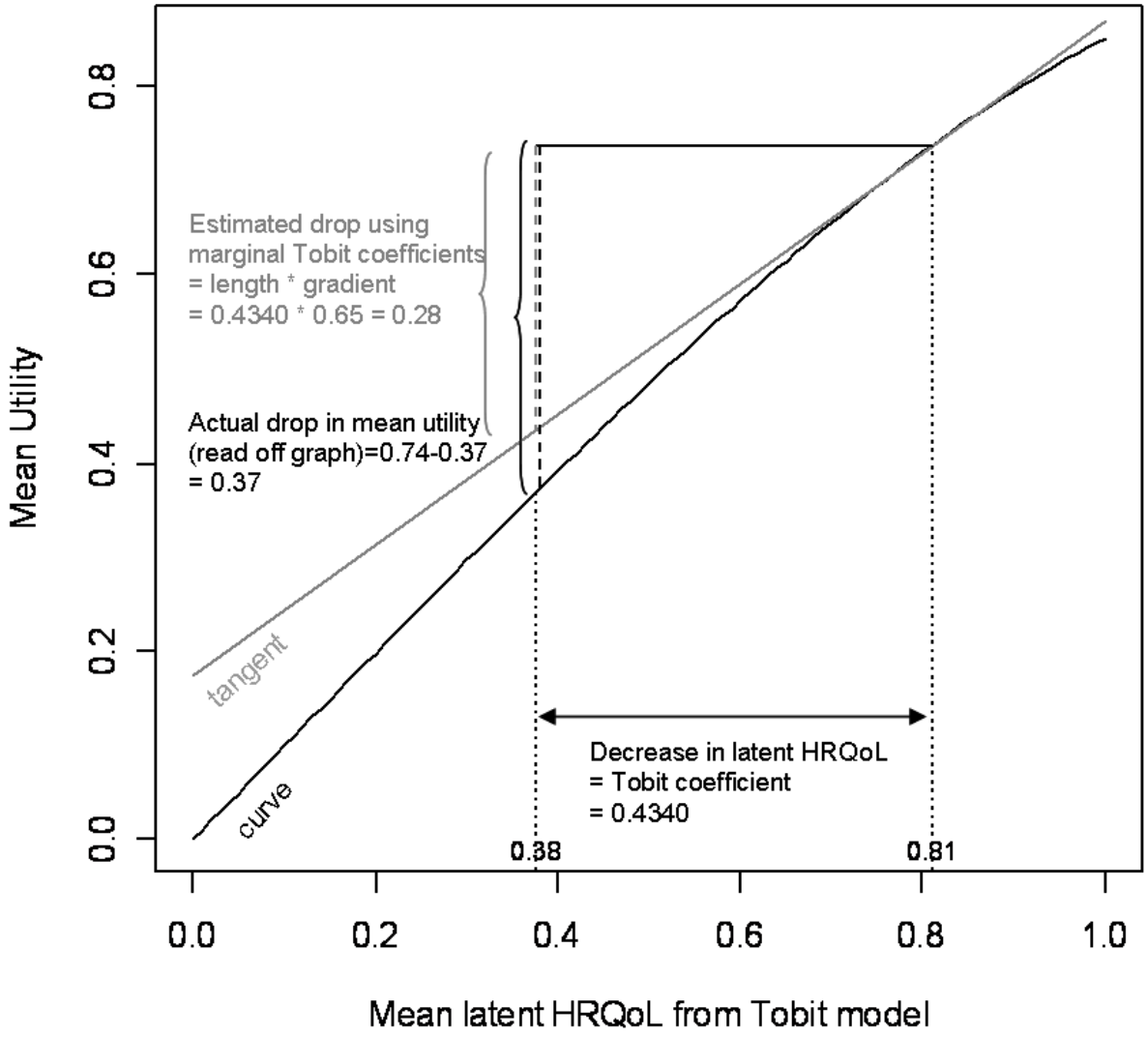
The Plot Thickens…

Marginal Coefficients

- Idea: Use Tobit or CLAD models to analyse a latent HRQoL
  - In 1-1 correspondence with utilities when <1
  - Is allowed to extend beyond 1
- But: we're not interested in the latent HRQoL
- So: Transform back to the original health utility scale

The Problem

- (sort of) OK so far
- Problem is the method used to transform back to the utility scale
  - Take regression coefficient and discount by the proportion of patients with utilities of 1
  - E.g. the regression coefficient for amputation in UKPDS is 0.43 using the Tobit model
  - 35% of the sample had a utility of 1
  - The discounted coefficient is 0.28 (0.43*0.65)
  - The actual coefficient is 0.37 (out by 24%)


Mean Utility
0.0
0.2
0.4
0.6
0.8
Estimated drop using
marginal Tobit coefficients
= length * gradient
= 0.4340 * 0.65 = 0.28
Actual drop in mean utility
(read off graph)=0.74-0.37
= 0.37
tangent
curve
Decrease in latent HRQoL
= Tobit coefficient
= 0.4340
0.38
0.81
0.0
0.2
0.4
0.6
0.8
1.0
Mean latent HRQoL from Tobit model

How'd they come up with that?

$$Y^* \sim N(X\beta^*, \sigma^2)$$

$$Y = Y^* \wedge 1$$

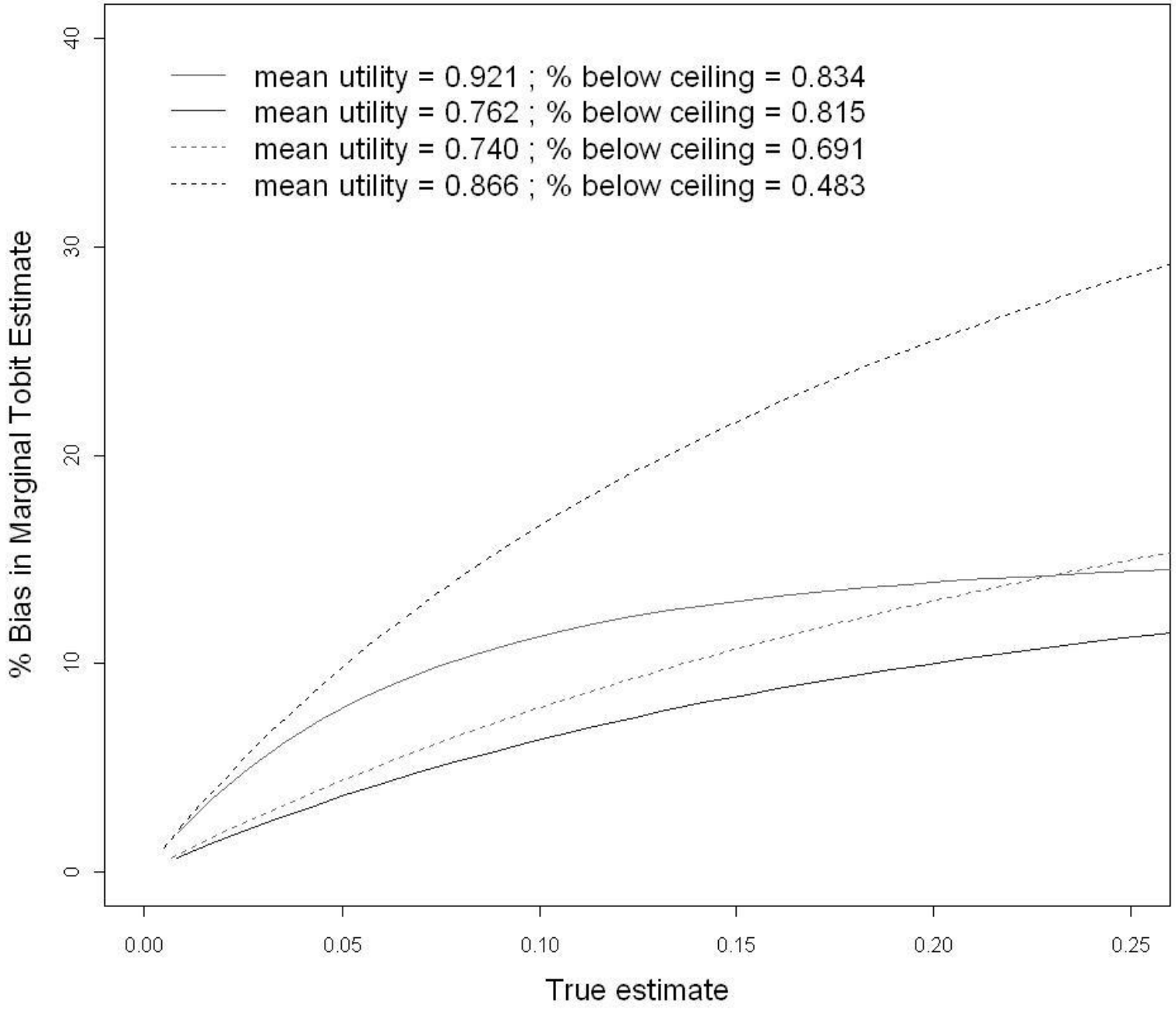
$$E(Y \mid X) = \frac{1}{\sigma}\int_{-\infty}^{1} y^* \phi\left(\frac{y^* - X\beta^*}{\sigma}\right) dy^* + 1 - \Phi\left(\frac{1 - X\beta^*}{\sigma}\right)$$

$$E(Y \mid X) = -\sigma\phi\left(\frac{1 - X\beta^*}{\sigma}\right) + X\beta^* \Phi\left(\frac{1 - X\beta^*}{\sigma}\right) + 1 - \Phi\left(\frac{1 - X\beta^*}{\sigma}\right)$$

$$\frac{\partial E(Y \mid X)}{\partial X_j} = \beta_j^* \Phi\left(\frac{1 - X\beta^*}{\sigma}\right)$$

Bias in published studies

- Clarke 2004 (diabetes):
  - Ranges from 10% (MI) to 24% (Amputation)
- Hahl 2006 (Type I diabetic complications):
  - Ranges from 6% (cardiovascular) to 21% (renal)
- Saarni 2007 (psychiatric disorders)
  - Ranges from 5% (alcohol dependence) to 28% (generalised anxiety disorder)
- *For most complications we looked at, bias was over 10%*


mean utility = 0.921 ; % below ceiling = 0.834
mean utility = 0.762 ; % below ceiling = 0.815
mean utility = 0.740 ; % below ceiling = 0.691
mean utility = 0.866 ; % below ceiling = 0.483
% Bias in Marginal Tobit Estimate
0
10
20
30
40
0.00
0.05
0.10
0.15
0.20
0.25
True estimate

Point #3

- Marginal Tobit & CLAD coefficients don’t make sense in this context

How did the mistake happen?

- Result relating derivatives of the censored mean to the uncensored mean appeared in the Economics literature (Greene et al.)
- First use of marginal Tobit/CLAD coefficients (Clarke et al.) cited this (Greene)
- Subsequent uses cited the Clarke et al.

Warning #2

- What people write in their **own** literature sometimes gets mis-interpreted
- Mis-interpretation gets perpetuated when use of methodology is justified by its use in past studies
- Can’t rely on peer-review to pick this up.

How should utilities be analysed?

Two-part models

Model

- Probability of hitting the ceiling
- Distn of utility given below ceiling

Two-part model: below the ceiling

- Without transformation

$$E(Y_i \mid Y_i < 1, X_i) = X_i\beta$$

- With transformation

$$E(\log(1 - Y_i) \mid Y_i < 1, X_i) = X_i\beta$$

- What is $E(Y|X)$?
  - Without transformation, $E(Y_i|X_i, Y_i<1)=X_i\beta$
  - With transformation, need distributional assumptions. If we assume $\log(1-Y)|(X,Y<1)\sim N(X_i\beta,\sigma^2)$, then $E(Y|Y<1,X)=1-\exp(X_i\beta+\sigma^2/2)$

Latent Class Models

If C_i is the latent class variable for individual i, with

$C_i \in \{1,2\}$, the latent class model would be

$$P(C_i = 2) = p_i$$

$$Y_i \sim N(X_i\beta_1, \sigma_1^2) \text{ if } C_i = 1$$

$$Y_i \sim N(X_i\beta_2, \sigma_2^2) \text{ if } C_i = 2$$

Then

$$E(Y_i \mid X_i) = (1-p_i)X_i\beta_1 + p_i X_i\beta_2$$

Linear regression

- Or just fit $E(Y|X)=X\beta$
- Use OLS
- To get std errors:
  - Robust standard errors
  - Non-parametric bootstrap
  - Do NOT use a semi-parametric bootstrap

Simulation

Same simulation set-up as before. Numbers are bias (bias/se)

Sample Size	OLS	TPM trans	TPM No trans	LCM
50	-0.00068	-7439.15	-0.00068	-0.00068
	(-0.34)	(-1.30)	(-0.34)	(-0.34)
100	0.00012	-29.92	0.00012	0.00014
	(0.11)	(-9.26)	(0.11)	(0.13)
200	-0.00170	-711.44	-0.00170	-0.00167
	(-1.77)	(-16.39)	(-1.77)	(-1.74)
500	-0.00030	-136.00	-0.00030	-0.00031
	(-0.47)	(-24.32)	(-0.47)	(-0.48)
1141	0.00049	-153.66	0.00049	0.00049
	(1.11)	(-43.21)	(1.11)	(1.12)

OLS: Coverage Probabilities

Sample size	Model-based	Robust	Semi-parametric bootstrap	Non-parametric bootstrap standard errors	Non-parametric bootstrap bca intervals
50	0.939	0.934	0.928	0.930	0.909
100	0.922	0.950	0.925	0.934	0.964
200	0.950	0.945	0.952	0.948	0.950
500	0.941	0.953	0.939	0.946	0.945
1141	0.927	0.944	0.925	0.940	0.94

Point #4

- OLS often does fine provided you account for heteroscedasticity

Comments from Reviewers

- Use of OLS when residuals are non-Normal will lead to biased estimates of regression coefficients

How do we teach regression?

- $Y_i=\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i$
- ε_i iid $\sim N(0,\sigma^2)$
- That is, residuals
  - are independent of one another
  - are Normally distributed
  - have a common standard deviation
- *We tend not to say what happens if the assumptions do not hold*

Warning #3

- Students leave our classes on linear regression believing that OLS is biased if Normality does not hold

Future Directions

- EQ5D probably most widely used generic measure
- Utilities are *interval* censored
- Measured utilities of 1 are over-estimates, on average
- Is there a way to incorporate the EQ-5D VAS in order to get a less biased estimate?

Summary: Utility Analyses

- Literature on analysis of health utility data very confusing
- Tobit & CLAD models not usually appropriate for economic analysis
- OLS will often work well
  - Use robust std errors or a non-parametric bootstrap
- If linearity is a problem, consider a GAM

Summary: General points

- Communication breakdown!
- We have our uses even in fields where PIs feel happy doing their own analysis
- Requires more than a cursory glance over methodology
- Do we over-simplify in our teaching?

Your thoughts?

Analyzing Health Utility data with Generalized additive models

Hoi Suen Wong

Department of Mathematics and Statistics
McMaster University, Hamilton

Feb 11, 2011

Analyzing Diabetes Hamilton data with OLS

Assumption regarding OLS

- $Cov(\epsilon_i, \epsilon_j) = 0$
- homogeneity of error terms
- no patterns observed when residuals plotted against the predicted values

Assumption for inferences of the parameters

- $\epsilon_i \sim N(0, \sigma^2)$

Figure 1: Residuals against the predicted values.

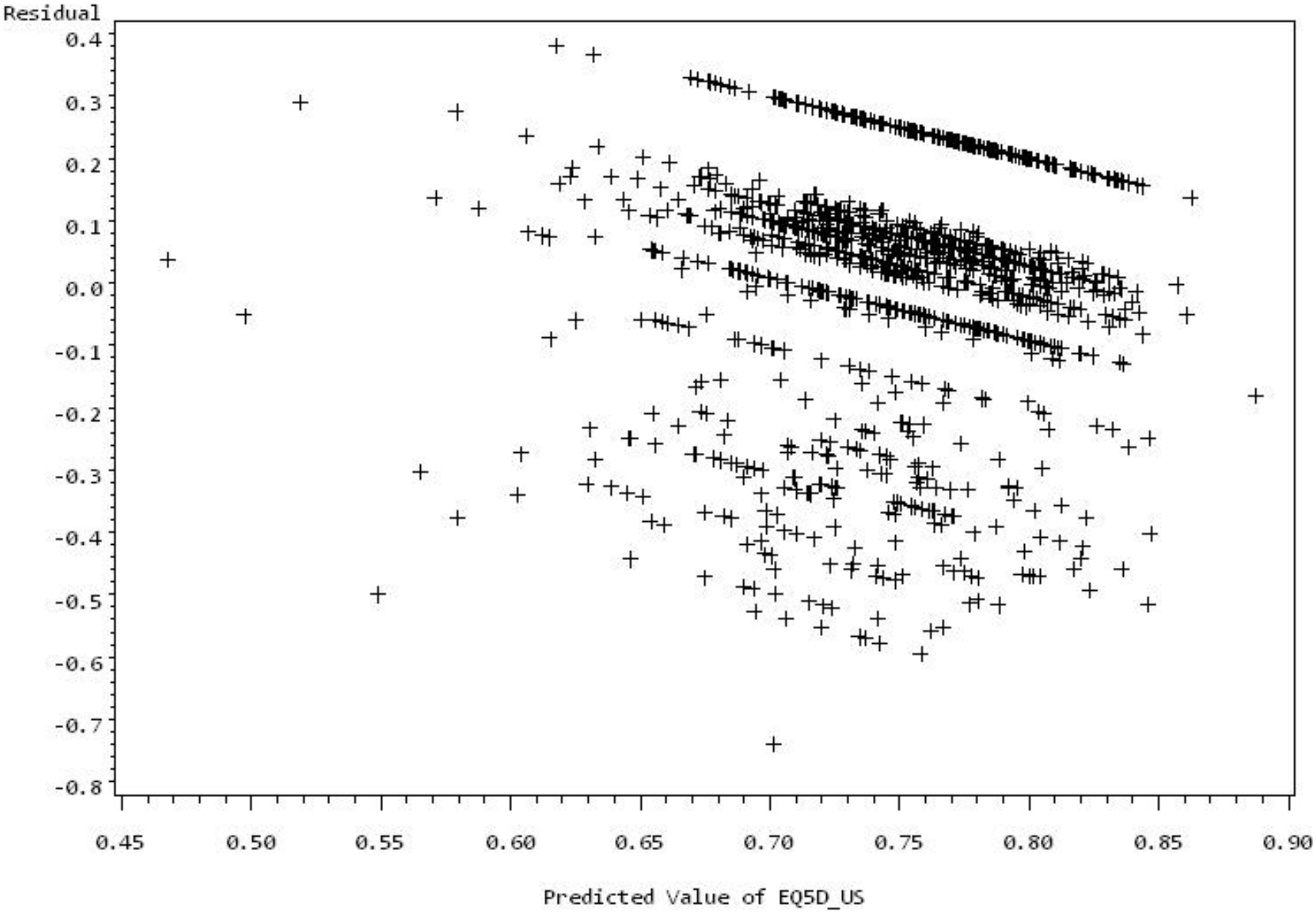


Figure 2: Residuals with the use of loess smoothing.

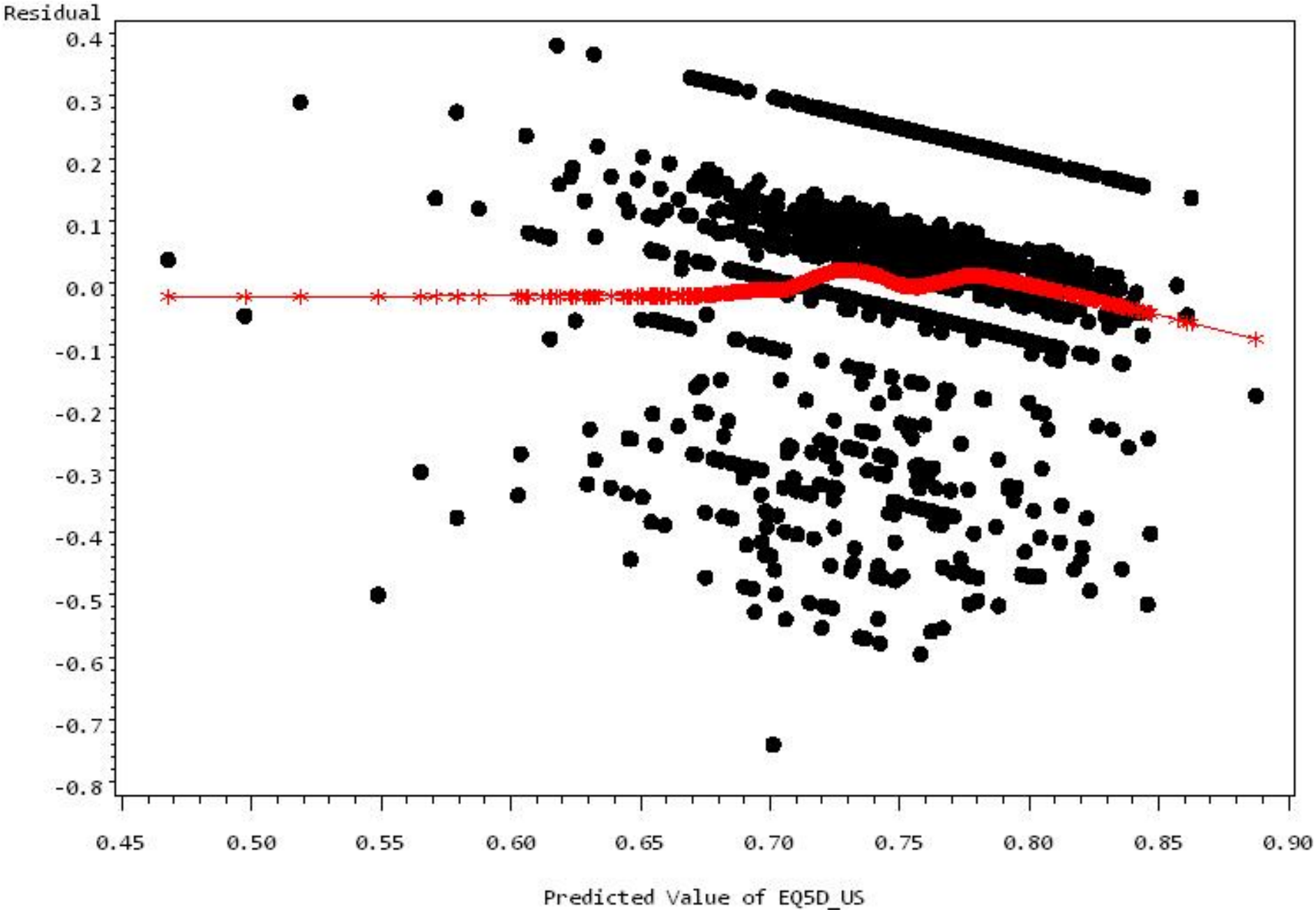


Figure 3: Histogram of the residuals.

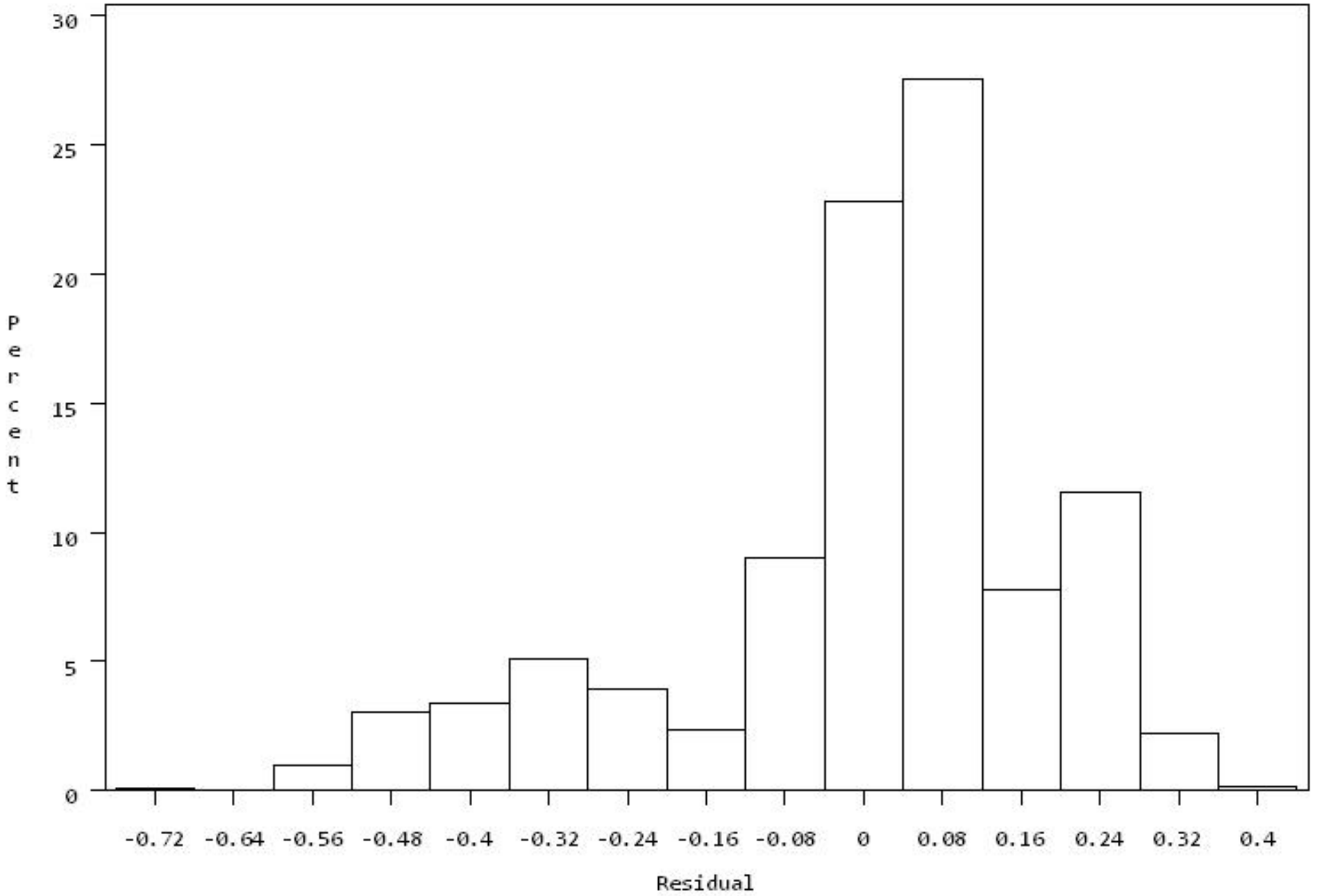


Table: Estimates of parameters and their standard error and p-value with the use of regression analysis.

Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t value*	*P-value*
Intercept	1	0.5170	0.0382	13.53	$<$0.0001
Foot Leg Amputation	1	-0.0631	0.0526	-1.2	0.2302
Stroke	1	-0.0462	0.0228	-2.03	0.0426
Heart Attack	1	-0.0586	0.0173	-3.39	0.0007
Kidney Failure	1	-0.1018	0.0378	-2.7	0.0071
Age	1	0.0029	0.0006	5.2	$<$0.0001
Gender	1	0.0515	0.0116	4.44	$<$0.0001
Duration of Diabetes	1	-0.0015	0.0006	-2.47	0.0136

Heteroscedasticity of error terms

Heteroscedasticity of error terms can be handled by using:

- Robust standard errors
- Bootstrapping

Robust standard errors and bootstrapping standard errors

Table: Robust standard errors, robust p-value, standard deviation, bootstrap p-values of each parameters estimates

Variable	*Parameter Estimate*	*Std. Error*	*Robust Standard Error*	*Bootstrap Standard Error*	*p-value*	*Robust p-value*	*Bootstrap p-value*
Intercept	0.0517	0.0382	0.0399	0.0006	<0.0001	0	0
Foot Leg Amputation	-0.0631	0.0526	0.0567	0.0513	0.2302	0.2302	0.2819
Stroke	-0.0462	0.0228	0.0237	0.0233	0.0426	0.0515	0.0555
Heart Attack	-0.0586	0.0173	0.0171	0.0179	0.0007	0.0006	0.0004
Kidney Failure	-0.1018	0.0378	0.0467	0.0480	0.0071	0.0295	0.0347
Age	0.0029	0.0006	0.0006	0.0006	<0.0001	0	0
Gender	0.0515	0.0116	0.0117	0.0113	<0.0001	0.0001	0.0001
Duration of Diabetes	-0.0015	0.0006	0.0006	0.0006	0.0136	0.0105	0.0128

Figure 4: Residuals against the parameter age with the use of loess curve.

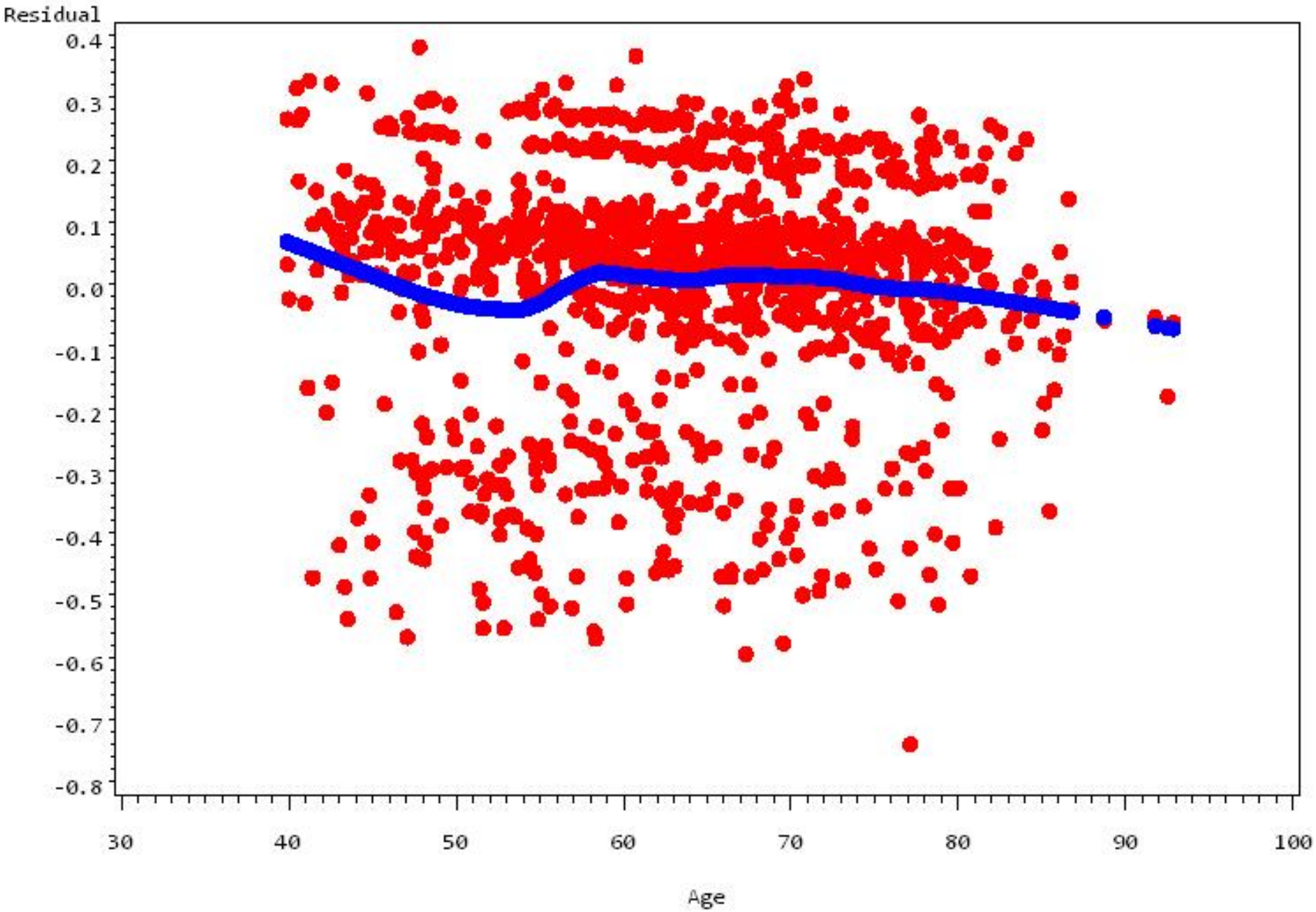
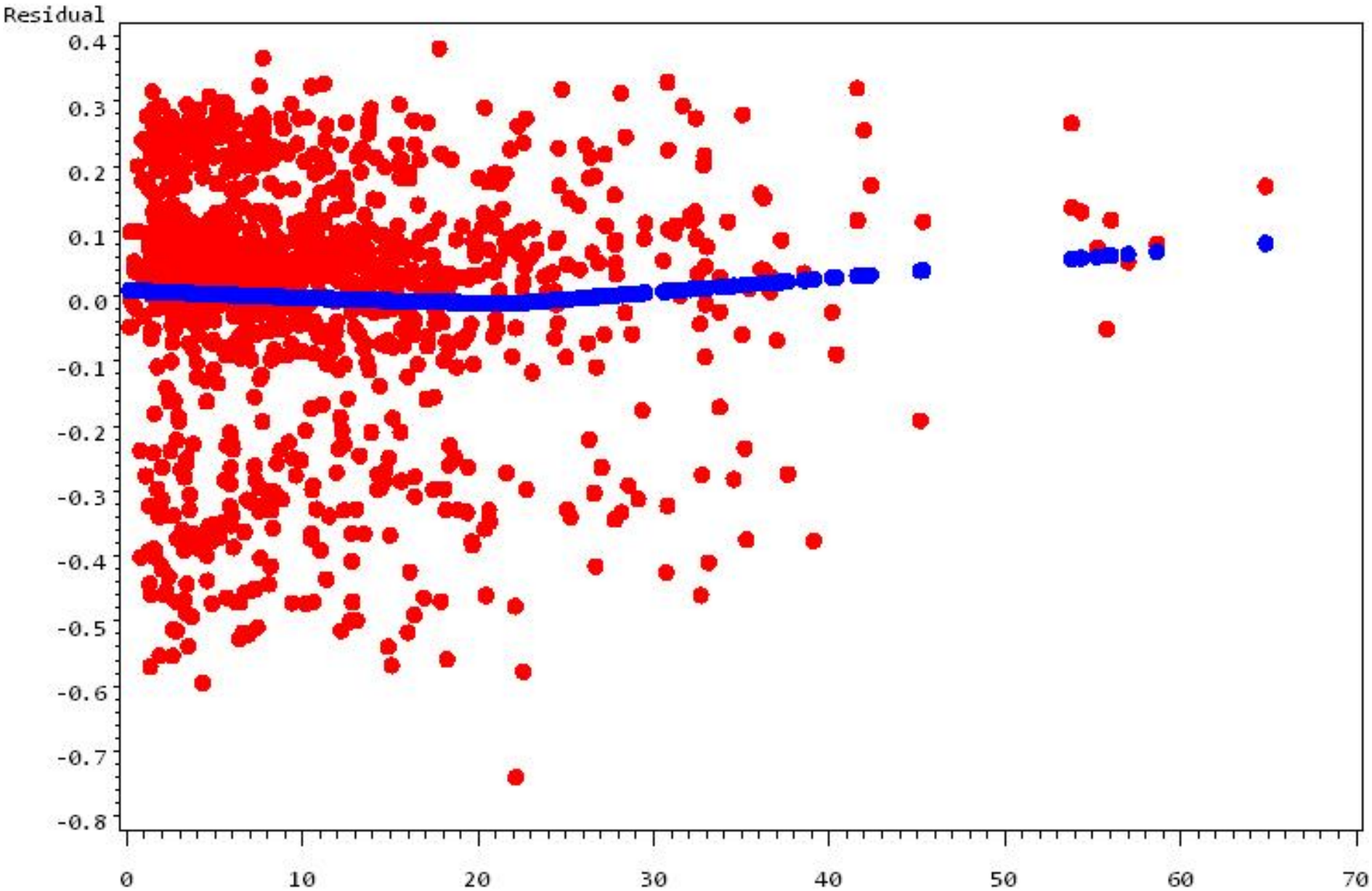


Figure 5: Residuals against the parameter duration of diabetes with the use of loess curve.



Ordinary least squares & Generalized additive models

1. OLS assumes the form:
   $\mu = E(Y_i \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$,
   $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2 \ \forall i$
2. GAM assumes the form: $Y = s_0 + s_1(x_1) + s_x(x_2) + ... + s_p(x_p) + \epsilon$
   $Cov(X_i, X_j) = 0$, $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2 \ \forall i, j$
   where $s_1(x_1), ..., s_p(x_p)$ are some arbitrary functions. The s_i can also be of more than one variable form, for example s_i may be $s(x_1, x_3)$. It also assumes the response variable is from the exponential family with possibility of different link functions.

Generalized additive models

- Model: $Y = s_0 + s_1(x_1) + s_x(x_2) + ... + s_p(x_p) + \epsilon$
  $Cov(X_i, X_j) = 0$, $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2 \ \forall i, j$
- possibility of different links: for example a binary variable with a logit link can be represented as $log(\frac{\mu}{1-\mu}) = s_0 + s_1(x_1) + ... + s_p(x_p) + \epsilon$
- The estimation of $s_0, s_1(x_1), s_x(x_2), ..., s_p(x_p)$ is done using a procedure called local scoring where a backfitting algorithm is employed.

Analysis of Diabetes Hamilton data with Generalized additive model

Table: Estimates of parameters and their standard error and p-value with the use of Generalized additive models with spline for the parameters age and duration of diabetes.

Variable	*Parameter Estimate (OLS)*	*Parameter Estimate (GAM)*	*Standard Error (GAM)*	*t value (GAM)*	*P-value (GAM)*
Intercept	0.517	0.510	0.0380	13.42	$<$0.0001
Foot Leg Amputation	-0.063	-0.060	0.0522	-1.15	0.2487
Stroke	-0.046	-0.047	0.0230	-2.06	0.0397
Heart Attack	-0.059	-0.056	0.0172	-3.28	0.0011
Kidney Failure	-0.102	-0.104	0.0376	-2.78	0.0055
Linear (Age)	0.003	0.003	0.0006	5.44	$<$0.0001
Gender	0.052	0.052	0.0115	4.52	$<$0.0001
Linear (Duration of Diabetes)	-0.002	-0.002	0.0006	-2.63	0.0087

Simulation

1. Simulation Model 1: Beta model with a lump mass at 1
   Purpose:
   - To assess the performance of GAM in estimating the marginal effect of heart attack after adjusting for a continuous variable.
   - To compare between GAM and OLS based on the bias and coverage probability given by each method.
2. Simulation Model 2: Two part Logarithmic Model (A more realistic model)
   Purpose:
   - To assess the performance of GAM in estimating the marginal effect of heart attack after adjusting for the person's age.
   - To compare between GAM and OLS based on the bias and coverage probability given by each method.

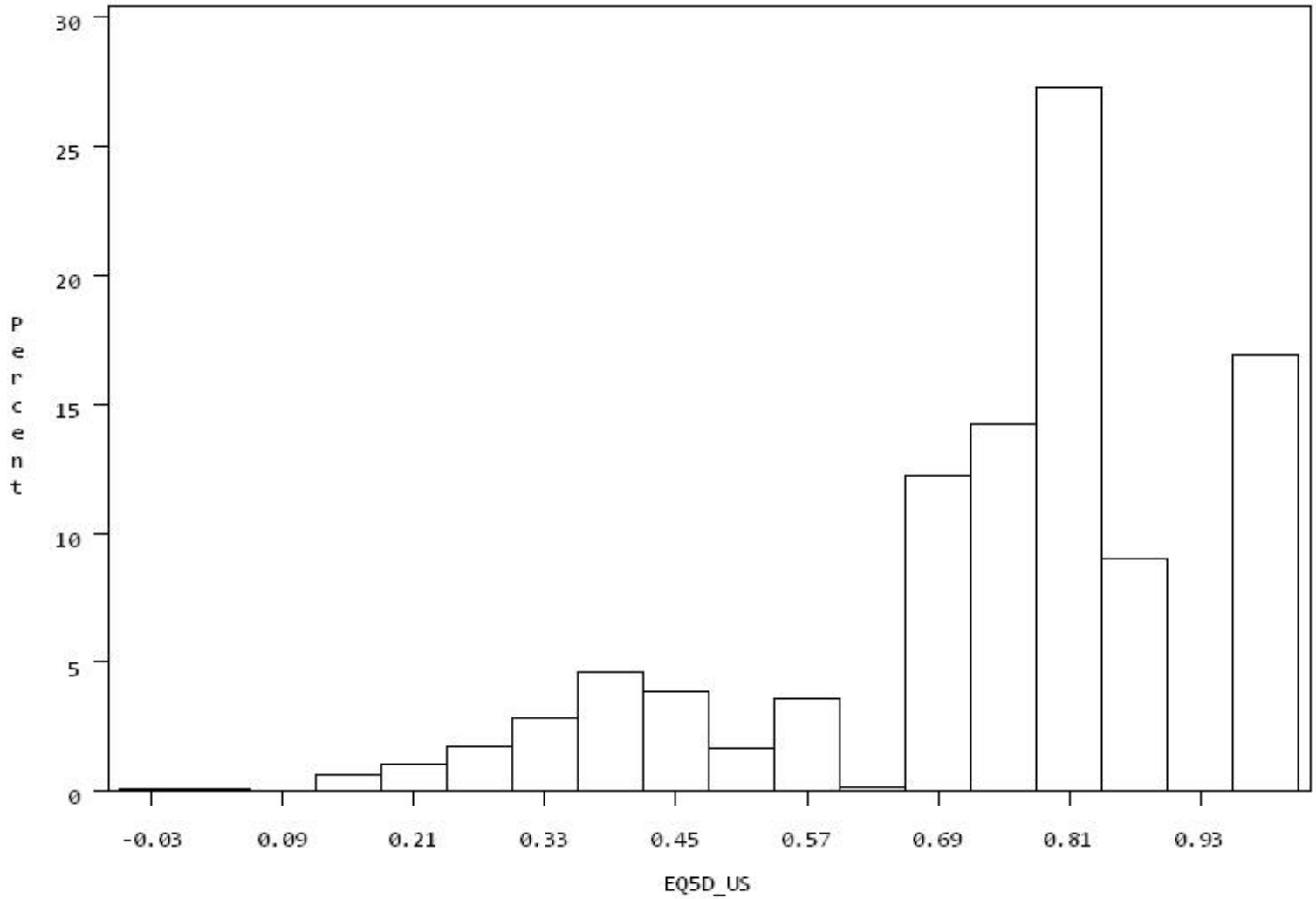


Figure 6: Histogram of the EQ5D value of the Diabetes Hamilton data set

Simulation Model 1: Beta model with a lump mass at 1

$$\begin{cases} X_1 \sim Uniform(0,1), X_2 \sim Bernoulli(p), logit(p) = logor * X_1 \\ Y = 1 * Ceiling + Y_{beta} * (1 - Ceiling), Ceiling \sim Bernoulli(q) \\ q = \frac{1}{1+exp(1+\gamma_1 X_1+\gamma_2 X_2)} \\ Y_{beta} \sim Beta(a,b), a = exp(2), b = exp(\alpha_1 X_1 + \alpha_2 X_2) \end{cases} \quad (1)$$

Where in the model, X_1 is taken to be a continuous variable, and X_2 is taken to be one of the most frequent complications in the data i.e. heart attack.

Two scenarios are investigated:

- α_2=2, γ_2=0.2
- α_2=0, γ_2=0

The true value of the marginal effect of heart attack

The expected value of Y from (1) is given by

$$E(Y|X) = \frac{1}{1+e^{1+\gamma_1 X_1+\gamma_2 X_2}}\left(1+\frac{e^{1+\gamma_1 X_1+\gamma_2 X_2}}{1+e^{-2+\alpha_1 X_1+\alpha_2 X_2}}\right) \tag{2}$$

So the marginal effect of heart attack is given by

$$E_{X_1}(\mu(X_1,1)-\mu(X_1,0)|X_2=1) \tag{3}$$

To calculate the true value:

- substitute X_2=1 into the equation (2) and integrate the equation from X_1=0 to X_1=1.
- substitute X_2=0 into the same equation and integrate the equation from X_1=0 to X_1=1.
- take the difference between the two

Some 100000 simulated Y value from simulation model 1

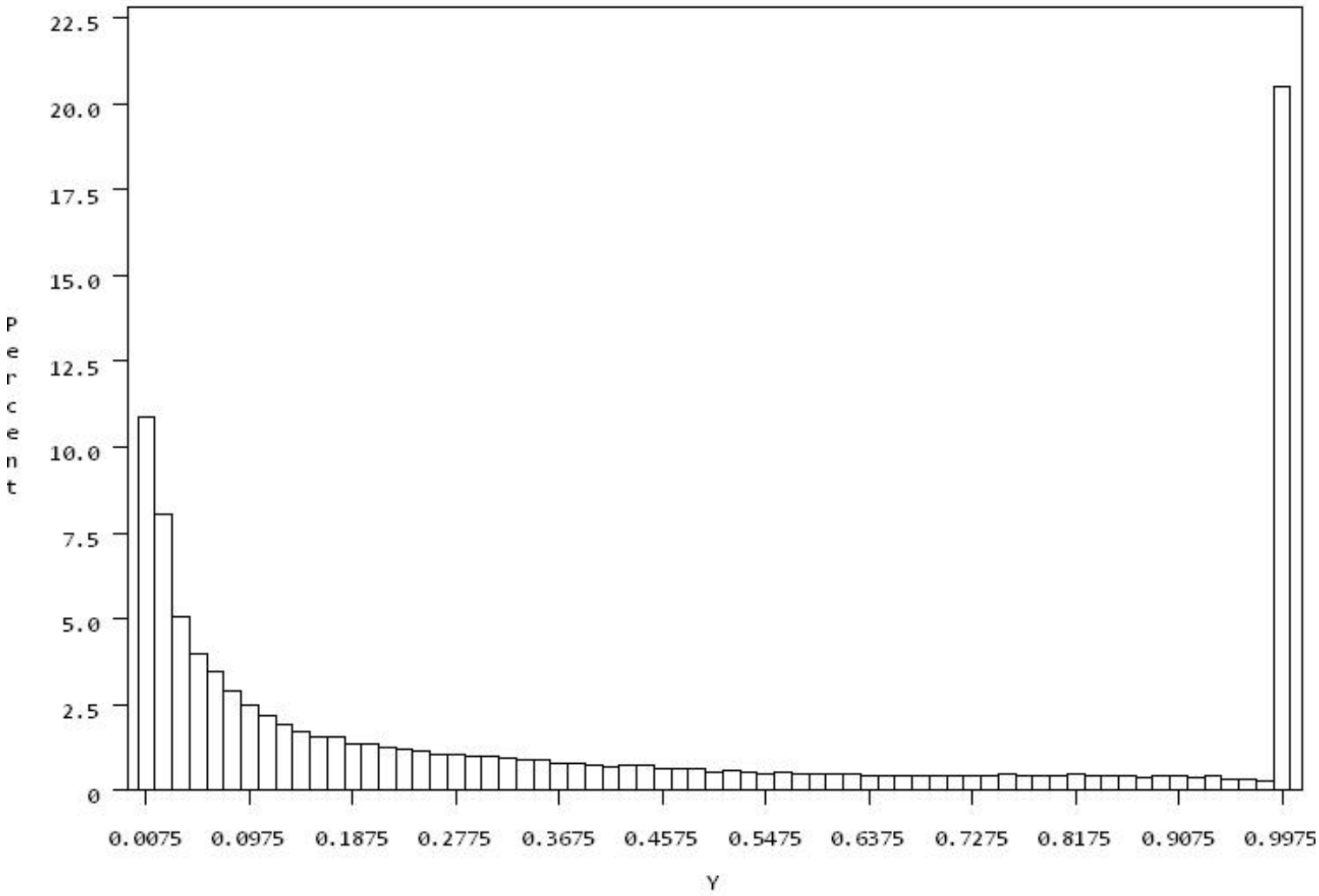


Figure 7: Histogram of the 100000 simulated value of Y from simulation model 1

Marginal effect of heart attack

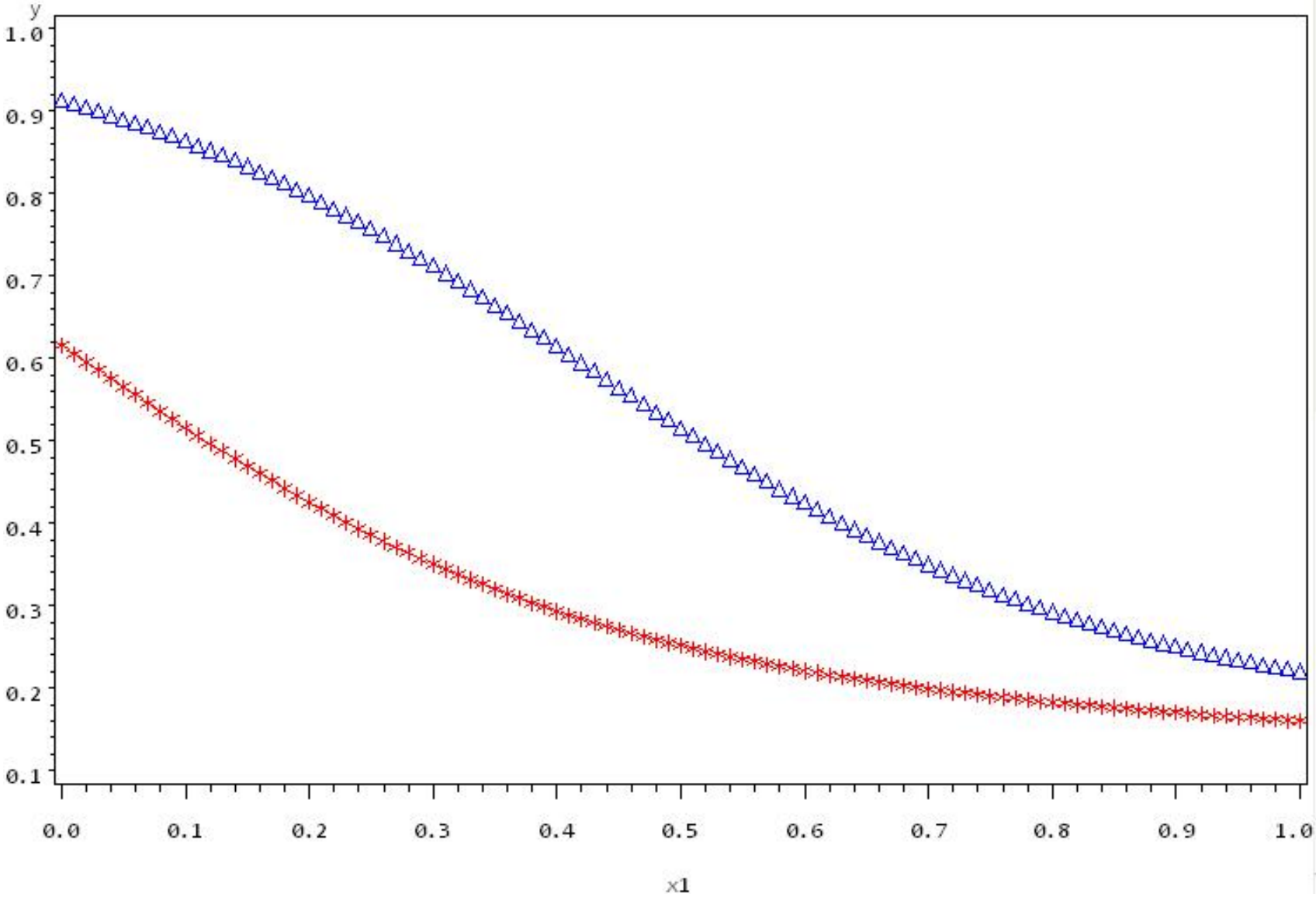


Figure 8: Change of expected value of the EQ5D score given the condition of heart attack (X_2=1, the lower curve) and no heart attack (X_2=0).

Simulation Results

Table: Comparison between OLS and GAM method for Simulation Model 1 with 1000 simulations and 5000 simulations when $\alpha_2 = 2, \gamma_2 = 0.2$

Model($\alpha_2 = 2, \gamma_2 = 0.2$)	Bias	[1] ESE	[2] ASE	[3] CP
	1000 simulations	*1000 simulations*	*1000 simulations*	*1000 simulations*
OLS no interaction	0.0093	0.0629	0.0677	0.955
GAM no interaction	0.0090	0.0635	0.0669	0.952
OLS with interaction	-0.0106	0.0679	N/A	N/A
GAM with interaction	-0.0034	0.0682	N/A	N/A
	5000 simulations	*5000 simulations*	*5000 simulations*	*5000 simulations*
OLS no interaction	0.0097	0.0643	0.0676	0.948
GAM no interaction	0.0096	0.0650	0.0668	0.945
OLS with interaction	-0.0108	0.0701	N/A	N/A
GAM with interaction	-0.0022	0.0700	N/A	N/A

[1] ESE stands for empirical standard errors

[2] ASE stands for average standard errors

[3] CP stands for coverage probability

Table: Comparison between OLS and GAM method for Simulation Model 1 with 1000 simulations and 5000 simulations when $\alpha_2 = 0, \gamma_2 = 0$

Model($\alpha_2 = 0, \gamma_2 = 0$)	Bias	ESE	ASE	CP
	1000 simulations	*1000 simulations*	*1000 simulations*	*1000 simulations*
OLS no interaction	0.00028	0.0559	0.0577	0.945
GAM no interaction	0.00023	0.0568	0.0571	0.950
	5000 simulations	*5000 simulations*	*5000 simulations*	*5000 simulations*
OLS no interaction	0.00024	0.0573	0.0578	0.950
GAM no interaction	0.00012	0.0577	0.0572	0.945

Simulation Model 2: Two part Logarithmic Model (A more realistic model)

$$\begin{cases} Y = 1 * Ceiling + Y_{lognormal} * (1 - Ceiling) \\ Ceiling \sim Bernoulli(q) \\ logit(q) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 \\ log(1 - Y_{lognormal}) \sim N(\beta_0 + \beta_1 X_1 + \beta_2 X_2, \sigma^2) \end{cases} \tag{4}$$

Where in the model, X_1 is taken to be the variable age, and X_2 is taken to be heart attack.

The true value of the marginal effect of heart attack

The expected value of Y from (4) is given by

$$E(Y|X) = q + (1-q)(1 - e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\sigma^2/2}) \tag{5}$$

where $q = \frac{e^{\alpha_0+\alpha_1 X_1+\alpha_2 X_2}}{1+e^{\alpha_0+\alpha_1 X_1+\alpha_2 X_2}}$.
So the marginal effect of heart attack is given by

$$E_{X_1}(\mu(X_1,1) - \mu(X_1,0)|X_2 = 1) \tag{6}$$

To calculate the true value:

- randomly select 100 people from the Diabetes Hamilton data set
- pick those who have heart attack value equal to one
- let X_2 to be one in the equation above, substitute the person's age, and calculate the equation's value
- let X_2 to be zero in the same equation, substitute the person's age, and calculate the equation's value
- calculate the difference for each chosen person, and then calculate the average of those differences

Some 100000 simulated Y values from Simulation Model 2

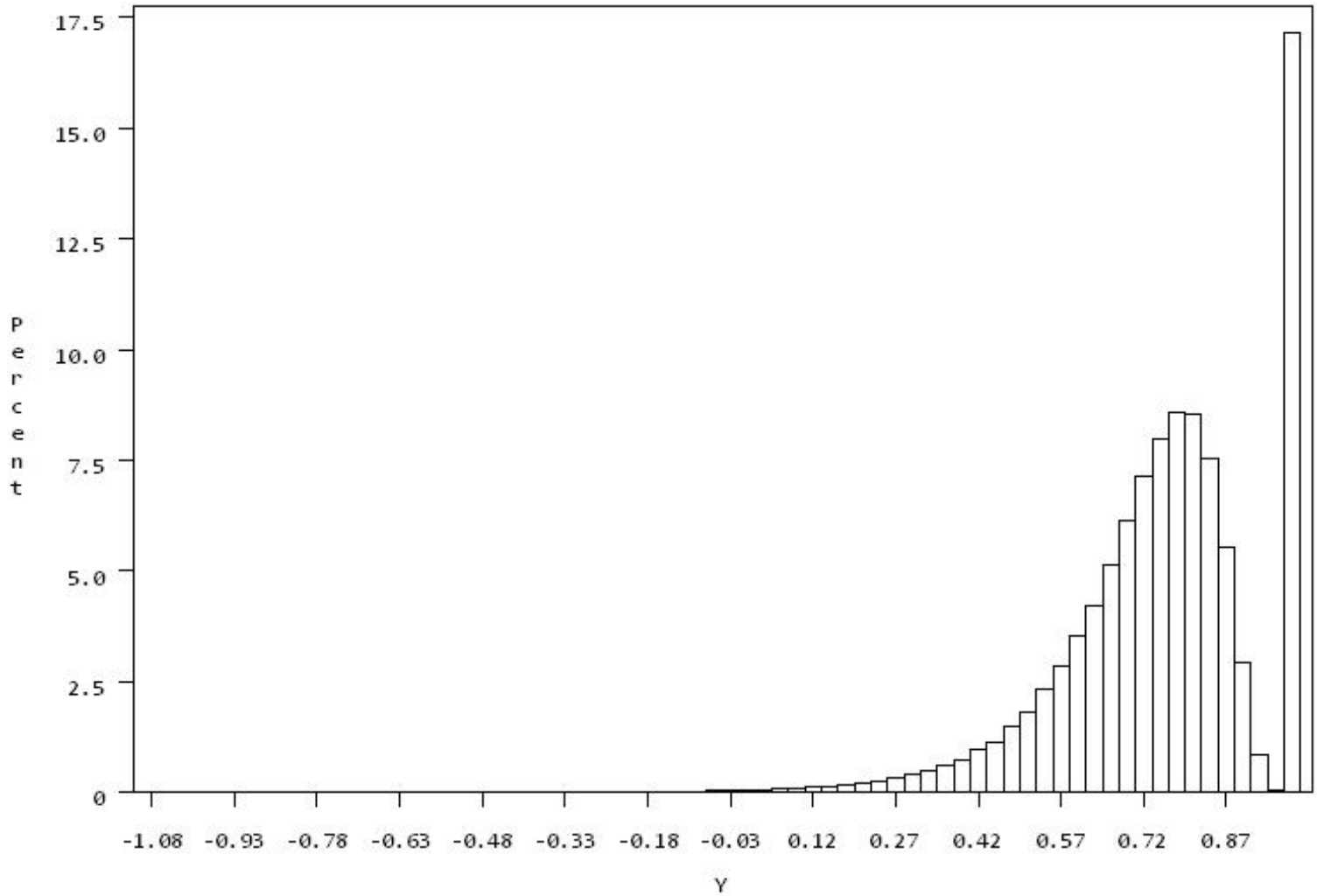
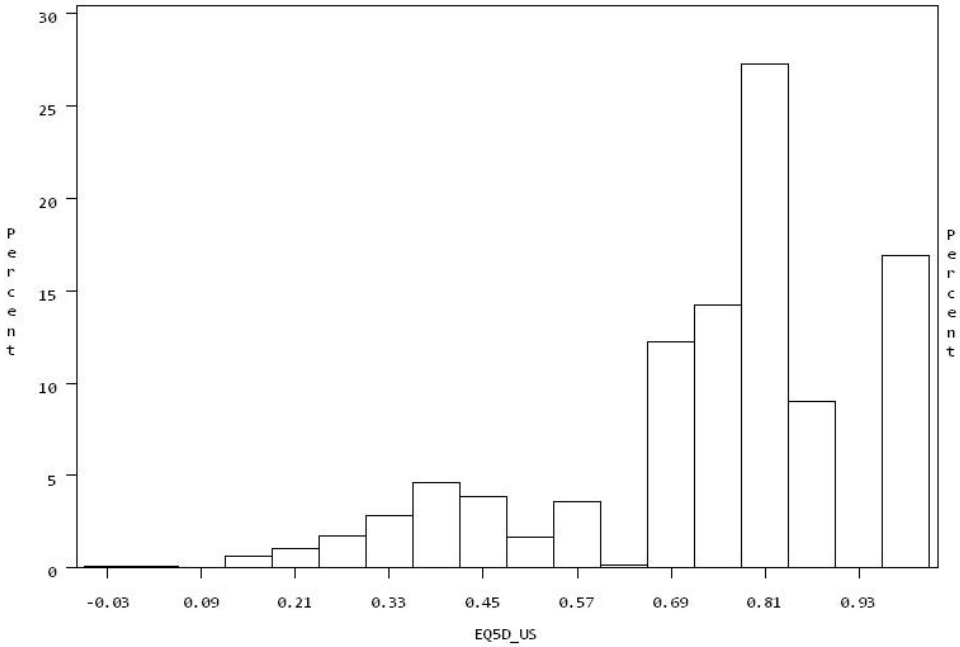
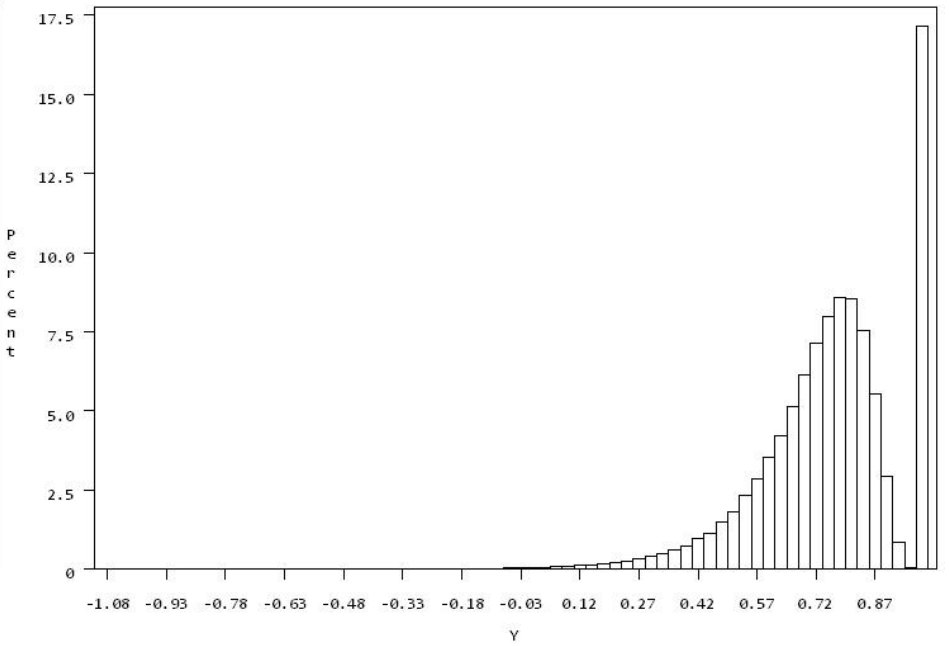


Figure 9: Histogram of the 100000 simulated value of Y from simulation model 2

Comparison between the real data set and Simulation Model 2





Marginal effect of heart attack

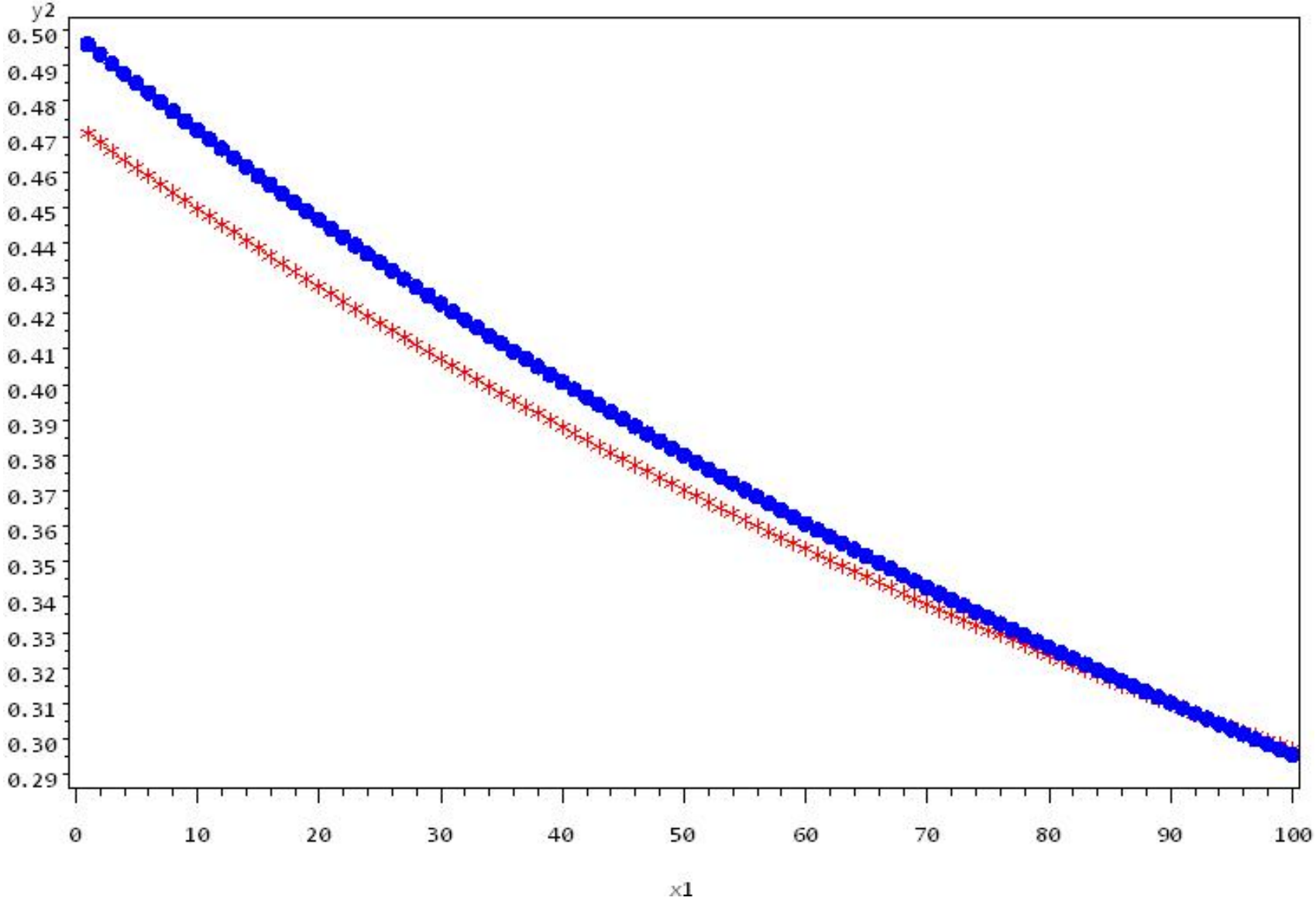


Figure 10: Change of expected value of the EQ5D score given the condition of heart attack (X_2=1) and no heart attack (X_2=0).

Simulations Results

Table: Comparison between OLS and GAM method for Simulation Model 2 with 1000 simulations and 5000 simulations

Model	Bias	ESE	ASE	CP
	1000 simulations	*1000 simulations*	*1000 simulations*	*1000 simulations*
OLS no interaction	0.00028	0.0559	0.0537	0.943
GAM no interaction	0.00027	0.0561	0.0534	0.941
OLS with interaction	0.00025	0.0560	N/A	N/A
GAM with interaction	0.00029	0.0562	N/A	N/A
	5000 simulations	*5000 simulations*	*5000 simulations*	*5000 simulations*
OLS no interaction	0.00028	0.0539	0.0538	0.948
GAM no interaction	0.00034	0.0540	0.0534	0.945
OLS with interaction	0.00027	0.0538	N/A	N/A
GAM with interaction	0.00035	0.0539	N/A	N/A

Conclusions

- The bias given by the GAM method is generally smaller than the OLS method from the result of Simulation 1
- The chance of making Type-1 error in hypothesis testing for the parameter is small for both GAM and OLS method from the result of Simulation 1.
- Both OLS and GAM methods produce small bias when applied to the Simulation 2 data.
- The coverage probability of each method are close to the expected value.
- Overal, GAM methods seem to be another good alternative method to analyzing health utility data.

Thank you